Advancing Artificial Intelligence Adoption in Oil and Gas Maintenance Through Advisory Recommendations

[1]Mohamad Afif bin Amir, [2]Mohammed F. Al-Rumaih, [3]Ahmed M. Alismail

[1]Senior Maintenance Engineer – Reliability, [2]Lead Business Systems Analyst, [3]Lead Business Systems Analyst,
[1,2,3]Corporate Maintenance Services Department, Saudi Aramco, Dhahran, Saudi Arabia
[1]mohamadafif.amir@aramco.com, [2]mohammed.rumaih.3@aramco.com, [3]ahmed.alismail@aramco.com

***Abstract*—Adopting artificial intelligence in the oil and gas industry requires attention to how intelligent systems interact with established workflows and professional judgment. This paper examines three approaches to maintenance notification classification: retrospective assessment, mandatory classification, and advisory recommendations. In the process considered, users select minor maintenance, normal maintenance, or corrective maintenance according to organizational criteria and operational judgment. Retrospective assessment identifies discrepancies after submission, whereas mandatory classification enforces the engine's selection. Advisory support presents contextual recommendations within the existing application and allows users to accept the advice or retain their original selection. Drawing on research in technology acceptance, task fit, automation, and human interaction with artificial intelligence, the paper argues that advisory recommendations offer the strongest fit for a process intentionally designed to accommodate discretion. Countervailing evidence on overreliance and human and artificial intelligence team performance defines the conditions under which this preference should be tested. A proposed framework connects recommendations, user responses, and independently assessed outcomes to support improvement. Effectiveness is evaluated through classification quality, user acceptance, workflow effort, and downstream rework. This paper does not present live plant test results. Instead, it offers a structured framework grounded in proven adoption models to guide future field deployment.**

***Index Terms*—Artificial intelligence, oil and gas, maintenance notifications, advisory recommendations, technology adoption, human oversight, decision support.**

I. INTRODUCTION

Artificial Intelligence (AI) tools add real value only when plant technicians and maintenance planners can smoothly apply the advice during routine work. Implementation must therefore address analytical capability and the conditions under which employees encounter, assess, and act on recommendations. A model's classification accuracy and the effectiveness of the overall implementation are related but distinct questions.

Experience at the plant shows clear user resistance when engineers must open a separate tool. In contrast, adoption improves when advice appears directly inside their everyday Computerized Maintenance Management System (CMMS). This observation has not yet been established through a controlled comparison. Research on information systems provides a plausible mechanism: switching costs contribute to resistance, while perceived value and organizational support can reduce it. The number of affected users describes the scale of implementation; the effort and uncertainty of changing working practices help explain individual resistance [1].

Maintenance notification creation provides a specific setting for examining this issue. In the process considered, users select minor maintenance, normal maintenance, or corrective maintenance. Selection remains flexible so that users can exercise judgment and account for operational circumstances. These categories are organization-specific, and their precise definitions must be established from internal maintenance procedures. They are not presented here as a universal or ordinal maintenance taxonomy.

This paper recommends advisory AI integrated into the existing notification process as the preferred implementation approach. It intervenes before submission, preserves the discretion built into the process, and captures feedback for evaluation. The contribution is a framework for comparing retrospective assessment, mandatory classification, and advisory recommendations, together with a testable explanation of why advisory support best fits this decision context.

II. EVIDENCE BASE AND SCOPE

The paper uses a targeted narrative synthesis of research on technology acceptance, user resistance, automation, human-AI interaction, and deployed machine-learning systems. Sources were selected for their relevance to implementation timing, workflow fit, decision authority, feedback, and evaluation. Supporting and countervailing findings are included. This is not an exhaustive systematic review, and no pooled effect or new empirical result is claimed.

The evidence supports mechanisms and design choices rather than a directly measured advantage in oil and gas maintenance notification creation. The adoption argument is informed by foundational information systems studies; experimental human-AI studies inform interaction design; workplace research illustrates measurable assistance; and a supplier example demonstrates an industry integration pattern. The proposed maintenance workflow and evaluation measures are the authors' synthesis. Their effectiveness must be established locally.

III. MAINTENANCE NOTIFICATION CLASSIFICATION AS A DECISION PROCESS

Selecting a notification type involves interpreting a maintenance requirement against organizational criteria. The AI engine may use the notification description, structured fields, and relevant equipment or maintenance records where available and authorized. Information available at the decision time should be distinguished from later findings that can support evaluation but would not have been available during creation.

The recorded information may not fully represent the circumstances known to the person creating the notification. A user may have additional knowledge about the observed condition, ongoing work, or the interpretation of procedures. Whether or not this information actually makes a difference in the classification is up to the organization's rules.

The implementation needs to differentiate between mandatory requirements and discretionary classifications. Deterministic validation can enforce required fields and unambiguous process rules. For the discretionary selection considered here, AI should help users evaluate their choices. The objective is to improve classification quality and consistency while preserving an effective means of considering contextual information.

IV. Comparison of Implementation Approaches

Retrospective assessment

In retrospective assessment, a separate AI solution evaluates completed notifications and compares recommended types with the recorded selections. A dashboard or key performance indicator (KPI) summarizes discrepancies and identifies records for review. This can support quality assurance and reveal recurring patterns.

Its effect on a specific record depends on subsequent action. A reviewer must assess the finding, determine whether the model is correct, and arrange any necessary correction. The dashboard does not by itself assist the user at the original decision point. AI-user disagreement should initially be reported as a discrepancy rather than a confirmed error, since using the model as the reference standard would make evaluation circular. Retrospective monitoring is therefore a useful supporting capability, but an incomplete response to improving classification during creation.

Mandatory classification

In mandatory classification, the engine determines the notification type and enforces that choice. This creates consistency with the model's output, but such consistency does not independently establish correctness. When the engine misinterprets the description or lacks relevant context, an exception route is needed to prevent its mistake from becoming the final record.

Enforcement also changes the allocation of authority in a process deliberately designed to allow judgment. It therefore requires evidence that the classification rules, model performance, and exception handling justify restricting discretion. It may suit separately validated requirements, but is not the preferred initial strategy for the discretionary selection considered here.

Advisory recommendation

In advisory recommendation, the engine evaluates the user's selection before submission. When another type appears more suitable, the existing application presents the alternative with a concise reason. The user can adopt the recommendation or retain the original selection.

This is the preferred strategy because it combines an immediate opportunity to improve the record with a practical route for correcting either human or model mistakes. It also records the interaction without requiring a separate review exercise. Fig. 1 compares the intervention point and decision authority of the three approaches. These are design characteristics, not measured comparative results.

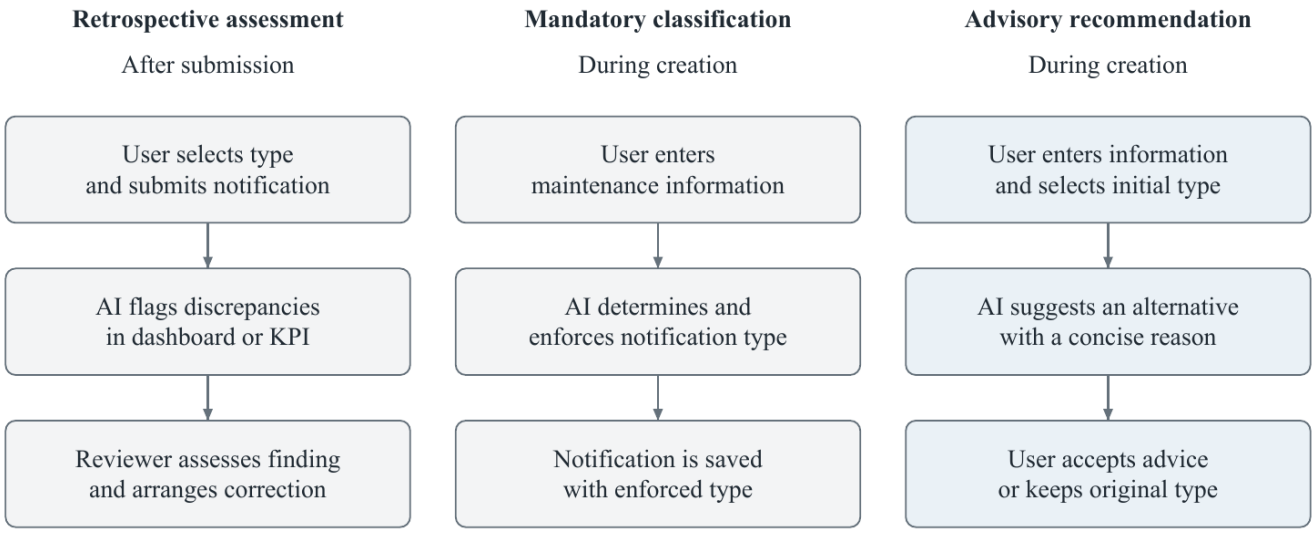

Fig. 1. Three approaches to maintenance notification classification. Advisory support combines intervention before submission with retained user choice. The comparison describes the proposed designs, not measured outcomes.

Interface location, intervention timing, and enforcement are separate design dimensions. A retrospective dashboard could be embedded in an existing application, and a standalone tool could offer timely advice. The strategies considered here reflect the organizational alternatives described above; a comparison among them evaluates complete implementation packages unless these dimensions are controlled separately.

V. Research Basis for the Advisory Preference

Adoption effort and fit with the task

Davis's Technology Acceptance Model identifies perceived usefulness and ease of use as central acceptance constructs [2]. Venkatesh and colleagues' Unified Theory of Acceptance and Use of Technology adds a broader organizational perspective through performance expectancy, effort expectancy, social influence, and facilitating conditions [3]. Applied here, integration should make

useful advice easy to access, while training and maintenance-team support remain necessary. Familiar software alone does not ensure adoption.

Goodhue and Thompson argue that performance benefits depend on both use and fit between the technology and the task [4]. A classification suggestion delivered beside the notification-type field fits the immediate selection task more directly than a later aggregate report. This is a design inference, not a result tested in their study. The advisory feature must therefore be judged by its relevance to notification creation, not by interaction counts alone.

Allocation of authority and user control

Parasuraman, Sheridan, and Wickens distinguish automation of information acquisition, analysis, decision selection, and action implementation [5]. This framework helps explain why automating analysis need not require automating the final choice. The proposed design delegates classification analysis to AI while retaining decision authority with the user. The framework supports deliberate function allocation rather than assuming that greater automation is necessarily preferable.

Dietvorst, Simmons, and Massey found that allowing people to modify algorithmic forecasts increased willingness to use an imperfect algorithm [6]. An acceptance-or-override mechanism applies the general principle of retained control to maintenance notification creation. Because the original studies concerned forecasting, the size of any adoption benefit in maintenance remains an empirical question. Nevertheless, the finding provides a direct psychological rationale for advisory support over unrestricted enforcement in a process that values discretion.

Timely assistance and contextual interaction

Kawamoto and colleagues' systematic review of clinical decision support identified workflow integration, actionable recommendations, and delivery at the time and location of decision-making among the features associated with improved practice [7]. This is evidence from healthcare rather than oil and gas, but supports presenting advice while a selection can still be changed. It provides a rationale for supplementing retrospective monitoring with immediate support.

Amershi and colleagues recommend contextually relevant information, efficient dismissal and correction, granular feedback, and cautious adaptation [8]. These principles map to an advisory feature that is easy to assess and decline within the existing task. They also support making the system's capabilities and limitations understandable. Background processing should reduce interaction effort while keeping the source of the advice transparent.

Appropriate reliance and countervailing evidence

Lee and See explain that trust influences reliance on automation and that its appropriateness depends on capabilities and context [9]. Advisory design makes selective reliance possible, but does not ensure it. Bansal and colleagues found that explanations increased acceptance of AI recommendations regardless of correctness and did not increase the complementary performance gains in their experiments [10]. Explanations should therefore be tested for their effect on decision quality rather than assumed to make recommendations trustworthy.

Buçinca, Malaya, and Gajos found that cognitive forcing interventions reduced overreliance relative to simple explainable-AI designs, but the most effective interventions received less favorable subjective ratings [11]. For the proposed process, obtaining an initial user selection before presenting advice is a reasonable design choice to test. Mandatory delays or burdensome justifications should not be introduced automatically; the trade-off between thoughtful review and workflow effort requires evaluation.

Vaccaro, Almaatouq, and Malone's meta-analysis of 106 experiments found that human-AI combinations were better than humans alone on average, but worse than the better of humans or AI alone. The synergy deficit was evident in decision tasks [12]. This is especially relevant because notification classification is a decision task. It prevents treating human oversight as proof of superior accuracy and supports evaluating AI alone as well as users alone and advisory assistance. The strong preference expressed here concerns suitability for the specified discretionary workflow, not demonstrated universal performance superiority.

VI. WORKPLACE AND OIL AND GAS RELEVANCE

Brynjolfsson, Li, and Raymond studied an AI conversational assistant among 5,172 customer-support agents and reported a 15 percent average increase in issues resolved per hour. Effects varied: less experienced and lower-skilled workers benefited more, while the most experienced and highest-skilled workers experienced small speed gains and small quality declines [13]. The study supports evaluating actual work outcomes and differences across user groups. Its effect size should not be transferred to maintenance classification, and it does not isolate embedded delivery from other implementation features.

Within oil and gas, SLB describes AI-assisted wellbore interpretation connected to Techlog and Petrel workflows, including human validation through interactive dashboards [14]. This illustrates the feasibility of combining AI processing, established engineering software, and professional review. As supplier documentation, it does not independently establish reduced resistance or superior adoption. It serves as an industry example rather than evidence of the proposed maintenance outcome.

VII. PROPOSED ADVISORY IMPLEMENTATION

Recommendation delivery

The user enters maintenance information and makes an initial notification-type selection. The AI engine evaluates the available information against approved organizational criteria. A recommendation is presented before submission only when the engine identifies an adequately supported alternative. The design can use a classification model, a rules-assisted model, or another validated analytical approach; advisory delivery does not require a particular AI architecture.

An illustrative message is: "Based on the information entered, [recommended type] may be more suitable than [selected type]. Reason: [brief explanation linked to the applicable classification criteria]." The available actions are "Use recommended type" and "Keep my selection." The bracketed fields describe the interface design and would be populated by the application.

The explanation should identify relevant input and the applicable criterion without claiming unsupported certainty. Acceptance updates the selected type. Retaining the original choice allows the process to continue and may offer an optional reason, such as missing context or incorrect interpretation. Reason categories should be developed with maintenance personnel. A legitimate override should not be treated as noncompliance merely because it differs from the AI output.

Interaction and integration controls

The interface should avoid repeatedly challenging an unchanged retained selection. Recommendations should be withheld when available information does not support a meaningful alternative. Any confidence threshold should be calibrated and validated against performance. If the advisory service is unavailable, the established notification process should remain usable, with the absence of AI review recorded separately from a user's rejection.

Integration should preserve application access controls, connect each recommendation to the correct notification and input version, and record when advice is presented. Recommendations should be refreshed or invalidated when relevant input changes. These are proposed engineering controls for avoiding stale advice and reconstructing decisions, rather than measured advantages of the implementation.

Figure 2 shows the proposed operating sequence, including the user's choice and the separate review and validation route for improvement. Retrospective reporting remains available across this sequence to support monitoring and targeted review.

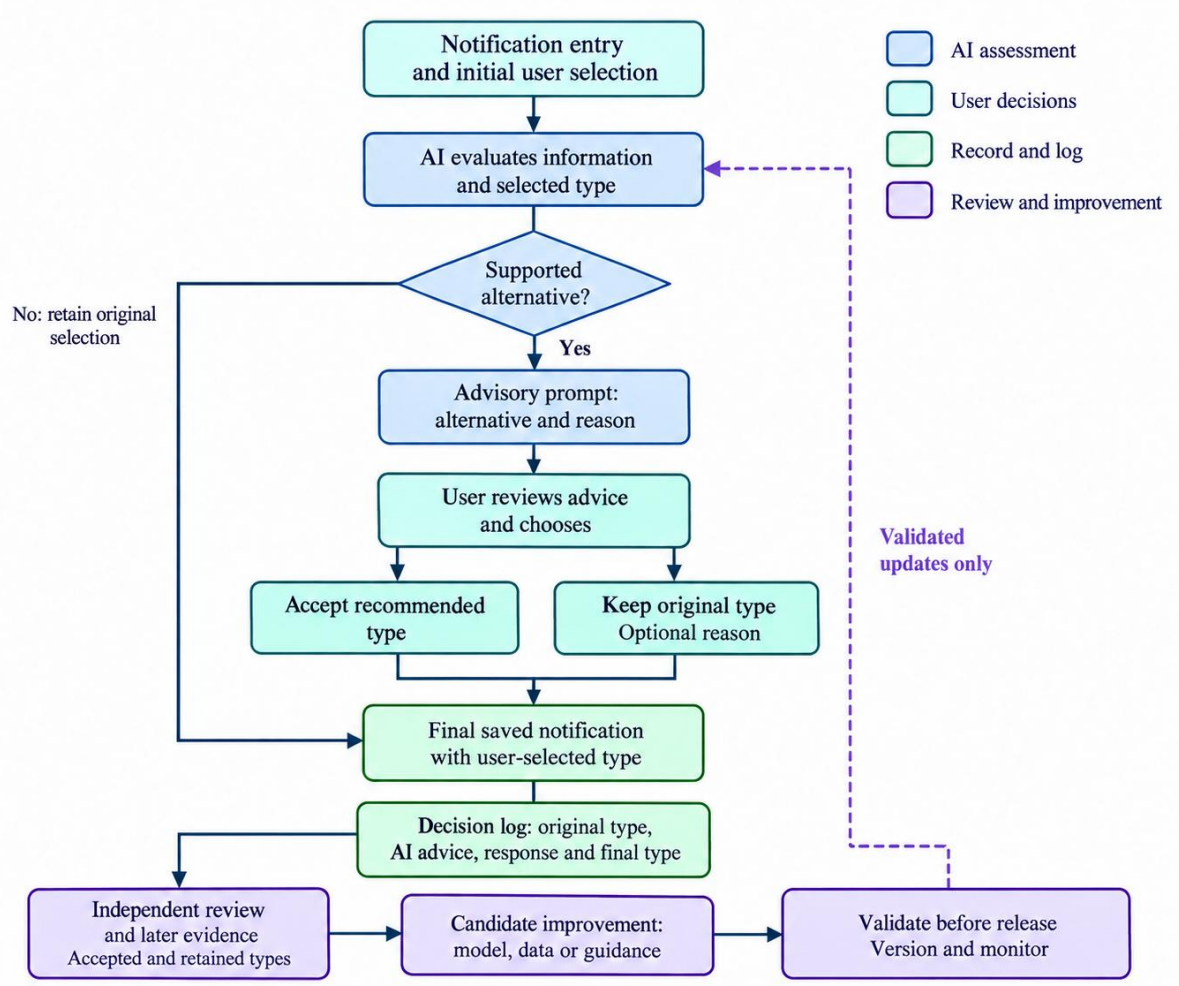

Fig. 2. Proposed advisory workflow and controlled improvement loop. Both accepted and retained selections are reviewed; acceptance alone is not a correct training label. Validated changes return to the deployed system.

VIII. FEEDBACK COLLECTION AND CONTROLLED IMPROVEMENT

Each recommendation event should preserve the original selection, relevant input snapshot, recommended type, explanation, model version, display event, user response, final saved type, and any optional reason. Subsequent expert assessment and classification corrections should be linked to the same event. A missing response, a recommendation that was never displayed, and an explicit rejection are different states.

Interaction records measure use and influence. Reviewed outcomes assess correctness. Neither acceptance nor rejection is automatically a reliable training label. Sculley and colleagues explain that deployed machine-learning systems can influence their own future training data and create feedback loops [15]. Indiscriminate training on accepted recommendations could therefore reinforce model errors.

The improvement process should combine interaction records with maintenance expert review and relevant later findings. Review should include representative accepted and overridden recommendations, as well as notifications receiving no advice, to detect errors the system failed to flag. Later information may support evaluation, but must not leak into training features intended to represent what was available at creation.

Disagreements may indicate model error, missing information, unclear rules, or inconsistent organizational practice. Improvements can therefore involve the engine, input fields, or classification guidance. Model updates should be evaluated on separate data before deployment, with version tracking and rollback. Changes in user behavior after an update should be assessed alongside technical performance.

IX. Evaluation framework

Outcomes and measurement

We expect built-in advisory prompts to drive cleaner maintenance records without adding screen time or taking authority away from maintenance planners. Classification quality should be the primary outcome, assessed against an independently reviewed reference. Acceptance and satisfaction are complementary measures, not substitutes for correctness. Table 1 sets out the proposed measures.

TABLE 1. PROPOSED EVALUATION MEASURES

Measure	**Proposed assessment**
Classification quality	Final type compared with independently reviewed reference classification.
Recommendation quality	Proportion of reviewed AI alternatives assessed as appropriate.
Acceptance and final adoption	Acceptance among displayed recommendations; final saved type reported separately.
Override quality	Proportion of reviewed retained selections assessed as justified.
Creation effort	Completion time, extra interactions, and frequency of advisory prompts.
Downstream rework	Subsequent classification corrections and associated review effort.
User experience and sustained use	Perceived usefulness, effort, control, and engagement over time.

Acceptance should use a declared denominator, such as accepted alternatives divided by distinct recommendations displayed during the observation period. Verified viewing can be reported separately if measured reliably. Final adoption should also be distinguished from clicking accept and subsequently changing the type. Results should be reported by notification category and user experience where sample sizes permit, so that frequent categories or large user groups do not conceal poor performance elsewhere.

Reference classifications should use documented criteria. Where feasible, reviewers should be blinded to the AI recommendation and implementation strategy, with disagreements adjudicated. Original-decision quality should be assessed using information available at creation; later findings should be analyzed separately to avoid judging users solely with hindsight. The evaluation should report uncertainty and unresolved cases rather than manufacture certainty from ambiguous rules.

Comparative study design

A comparison should hold the AI engine, available inputs, relevant training, and case mix as similar as practicable across retrospective, mandatory, and advisory strategies. A no-AI group helps identify improvement over current practice. An offline AI-alone assessment provides a model benchmark without requiring live enforcement before its suitability is established.

Random allocation by user, team, or site can be considered according to operational constraints and the risk of information sharing between groups. If random allocation is impractical, a phased comparison should account for differences in workload, equipment populations, experience, and time periods. The primary outcome, observation window, practically important improvement, and analysis approach should be specified in advance. Sample size should reflect the baseline error rate and any clustering.

Because the strategies change both timing and authority, an overall comparison evaluates implementation packages. Isolating the effect of user choice requires comparing advisory and enforced classification within otherwise similar workflows. Longer-term evaluation should examine sustained use and rework after initial novelty has diminished. Operational or financial benefits should be attributed to the intervention only when a credible comparison supports that inference.

X. Discussion

Advisory support is strongly preferred for the process considered because it improves the opportunity for correction at the point of creation while preserving intentionally permitted judgment. Retrospective assessment remains valuable for assurance, but depends on later action. Mandatory classification requires stronger justification for removing discretion and a reliable mechanism for exceptions.

The preference is conditional on recommendations being relevant, understandable, and sufficiently accurate to warrant attention. User choice can protect against some model errors, but can also preserve human mistakes or introduce inappropriate rejection. The countervailing literature therefore strengthens the evaluation requirements rather than being excluded from the argument. If local evidence demonstrates that another strategy yields better validated decisions with acceptable workflow and organizational consequences, the implementation choice should be reconsidered.

The framework's limitations are the observational nature of the motivating experience, the transfer of findings across professional domains, and the absence of comparative maintenance data. It does not establish a numerical adoption benefit, reduction in downtime, or return on investment. The proposed contribution is a reasoned starting design and a method for testing it. Clear criteria, useful advice, proportionate interaction, and independent review are necessary for translating that design into operational value.

XI. Conclusion

For maintenance tasks that rely on plant experience, advisory AI is the best starting point. It offers timely guidance without locking planners into rigid automated decisions. Presenting an alternative within the application before submission creates an immediate opportunity to improve the record while allowing a justified original selection to remain.

Research on switching costs, task fit, user control, workflow integration, and levels of automation supports this preference. Evidence on overreliance and human-AI performance establishes its limits: neither acceptance nor retained oversight proves correctness. Success should be demonstrated through validated classification quality, manageable effort, sustained engagement, and

reduced rework. Linking these measures to reviewed feedback offers a practical basis for improving both the AI engine and the maintenance process.

REFERENCES

[1] H.-W. Kim and A. Kankanhalli, "Investigating user resistance to information systems implementation: A status quo bias perspective," MIS Quarterly, vol. 33, no. 3, pp. 567–582, 2009. doi: 10.2307/20650309

[2] F. D. Davis, "Perceived usefulness, perceived ease of use, and user acceptance of information technology," MIS Quarterly, vol. 13, no. 3, pp. 319–340, 1989. doi: 10.2307/249008

[3] V. Venkatesh, M. G. Morris, G. B. Davis, and F. D. Davis, "User acceptance of information technology: Toward a unified view," MIS Quarterly, vol. 27, no. 3, pp. 425–478, 2003. doi: 10.2307/30036540

[4] D. L. Goodhue and R. L. Thompson, "Task-technology fit and individual performance," MIS Quarterly, vol. 19, no. 2, pp. 213–236, 1995. doi: 10.2307/249689

[5] R. Parasuraman, T. B. Sheridan, and C. D. Wickens, "A model for types and levels of human interaction with automation," IEEE Transactions on Systems, Man, and Cybernetics—Part A: Systems and Humans, vol. 30, no. 3, pp. 286–297, 2000. doi: 10.1109/3468.844354

[6] B. J. Dietvorst, J. P. Simmons, and C. Massey, "Overcoming algorithm aversion: People will use imperfect algorithms if they can (even slightly) modify them," Management Science, vol. 64, no. 3, pp. 1155–1170, 2018. doi: 10.1287/mnsc.2016.2643

[7] K. Kawamoto, C. A. Houlihan, E. A. Balas, and D. F. Lobach, "Improving clinical practice using clinical decision support systems: A systematic review of trials to identify features critical to success," BMJ, vol. 330, art. 765, 2005. doi: 10.1136/bmj.38398.500764.8F

[8] S. Amershi et al., "Guidelines for human-AI interaction," in Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, 2019, pp. 1–13. doi: 10.1145/3290605.3300233

[9] J. D. Lee and K. A. See, "Trust in automation: Designing for appropriate reliance," Human Factors, vol. 46, no. 1, pp. 50–80, 2004. doi: 10.1518/hfes.46.1.50_30392

[10] G. Bansal et al., "Does the whole exceed its parts? The effect of AI explanations on complementary team performance," in Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, 2021, pp. 1–16. doi: 10.1145/3411764.3445717

[11] Z. Buçinca, M. B. Malaya, and K. Z. Gajos, "To trust or to think: Cognitive forcing functions can reduce overreliance on AI in AI-assisted decision-making," Proceedings of the ACM on Human-Computer Interaction, vol. 5, no. CSCW1, art. 188, pp. 1–21, 2021. doi: 10.1145/3449287

[12] M. Vaccaro, A. Almaatouq, and T. Malone, "When combinations of humans and AI are useful: A systematic review and meta-analysis," Nature Human Behaviour, vol. 8, pp. 2293–2303, 2024. doi: 10.1038/s41562-024-02024-1

[13] E. Brynjolfsson, D. Li, and L. Raymond, "Generative AI at work," The Quarterly Journal of Economics, vol. 140, no. 2, pp. 889–942, 2025. doi: 10.1093/qje/qjae044

[14] SLB, "Wellbore interpretation," product documentation, accessed Sep. 11, 2026. Online: SLB product documentation

[15] D. Sculley et al., "Hidden technical debt in machine learning systems," in Advances in Neural Information Processing Systems 28, 2015. Online: conference paper